## TITLE OF THE INVENTION

MARKER AT THE VITAMIN D RECEPTOR GENE FOR DETERMINING BREAST CANCER SUSCEPTIBILITY

## FIELD OF THE INVENTION

The present invention relates to breast cancer. The invention further relates to a marker at the vitamin D receptor gene or equivalents thereof to prognose, diagnose or treat breast cancer. As well, the invention relates to a method for determining breast cancer susceptibility, prognosis and response to therapy based on a determination of a genotype at the vitamin D receptor locus or at a marker in linkage disequilibrium therewith. The invention further relates to screening assays to identify and select agents which can be used in the treatment of breast cancer.

## BACKGROUND OF THE INVENTION

Breast cancer is one of the most frequent cancer in women and causes a significant proportion of deaths by cancer as well. Two genes, BRCA1 and BRCA2, have been identified up to now as being associated with some familial forms of breast cancer. Together, these two genes account for 5% to 10% of all cases of breast cancer and possibly up to 70% of familial breast cancer cases.

Up to now, screening and diagnostic is normally carried out by physical breast examination first. A breast examination by a patient or a physician begins with a visual inspection for asymmetric breast size, nipple inversion, bulging, or dimpling. An underlying cancer is sometimes detected by having the patient press both hands against the hips or the palms together in front of the forehead. This contracts the pectoral muscles, and a subtle dimpling of the skin may appear if a Cooper's ligament has been entrapped by a growing tumor.

Breast cancers have the same clinical characteristics in older as in younger women. Cancer is usually suspected when changes are noted on mammography or when a breast lesion is seen or felt. Lesions usually can be felt as firm nodules within the breast. Ulcerations may occur, and lesions within or near the nipple may produce discharge Sometimes breast cancer is discovered only after metastatic lesions cause bone fractures, neurologic changes, hypercalcemia, liver failure, or ascites

When a tumor is detected by physical examination, bilateral mammograms are normally obtained to rule out occult lesions. Certain radiographic images, such as speckled calcifications or tissue infiltration, suggest cancer, while a cystic appearance suggests a benign process. Even an apparently benign finding on mammogram requires further evaluation. Generally the diagnosis is established by fine needle aspiration. Fine needle aspiration allows collection and cytological examination of cystic fluid and is helpful in planning definitive treatment of breast cancer. Although a positive result on fine needle aspiration is diagnostic, a negative result is usually followed by an open biopsy. Now a day, there is still no specific test for assaying predisposition or resistance to breast cancer.

Since the discovery of the human vitamin D receptor (VDR) gene, mutations in this gene have been associated with osteoporosis and with prostate cancer. As well, VDR gene polymorphisms studied (i.e. *BsmI* or *FokI*) have been associated with osteoporosis as well as prostate cancer (Morrison, N.A., et al., *Nature*, 367:284-287, 1994 and USP 5,593,033). As well, an association between the *BsmI* VDR polymorphism and metastatic breast cancer was reported (Ruggiero et al., 1998, Oncol Res 10 43-6). However, the Ruggiero et al study did not find any effect of VDR in general susceptibility to breast cancer but only an association of VDR genotype with the presence of metastases of breast cancer. Thus, a prediction of the general susceptibility of women to develop breast cancer in their lifetime, or of an association between

ART 34 AMDT

the VDR gene and breast cancer independently of metastases of breast cancer has yet to be reported.

Ingles et al. (Am. J. Human Genet., Vol. 61, p. A201 (1997) abstract) teaches an association between the vitamin D receptor (VDR) gene and breast cancer. Unfortunately, this association for breast cancer risk was identified with a FokI polymorphism in the 5' portion of VDR, only in one of the two cohorts of women tested, since hispanic women did not show such an association. There thus remains a need to identify an association between a combination of genotypes at VDR and breast cancer predisposition and, in particular, an association which can be identified in a larger proportion of patients as those of the prior art.

Yamagata et al. (Am. J. Human Genet., Vol. 61, p. A388 (1997) abstract) teaches an association with a single VDR marker (the BsmI in 3') and breast cancer in Japanese women. No combination of VDR loci is taught or suggested. There thus remains a need to identify an association between a combination of genotypes at VDR and breast cancer predisposition and, in particular, an association which can be identified in a larger proportion of patients as those of the prior art.

There thus remains a need to provide a genetic assay for determining the predisposition and/or resistance to breast cancer, development of breast cancer and responsiveness to therapeutic modalities.

While some markers have been identified as genetic determinants for breast cancer and/or as risk factors to develop same (i.e. BRCA1 and BRCA2), there remains a need to identify new markers therefor. More specifically, there remains a need to provide means to determine a predisposition to breast cancer and/or responsiveness to therapy to breast cancer, by analyzing allelic variations in genes associated with breast cancer. In addition there remains a need to identify patients who are likely to benefit from a particular prevention or therapeutic treatment program. Further, there remains

ART 34 AMDT

a need to provide assays to screen for compounds (i.e. hormones, molecules acting on hormone receptors or other agents) that could be beneficial to patients.

The present invention seeks to meet these and other needs.

The present description refers to a number of documents, the content of which is herein incorporated by reference, in their entirety.

## SUMMARY OF THE INVENTION

One aim of the present invention is to provide a genetic assay for determining the predisposition to breast cancer and/or response to breast cancer treatment.

In one particular embodiment, the determination of a predisposition to breast cancer is assessed by a determination of the genotype of the VDR gene, at at least one locus thereof, or at a locus in linkage disequilibrium therewith. In a preferred embodiment a determination of the VDR genotype, comprises a determination of the *BsmI* polymorphism, wherein certain combinations of VDR gene biallelic *BsmI* polymorphism (bb genotype; presence

of the restriction site on both alleles) have a significant but slightly increased risk of developing breast cancer as compared to the category with the smallest risk (BB genotype; absence of the restriction site on both alleles).

In another embodiment, the assessment of a predisposition and/or protection to breast cancer comprises a determination of the VDR genotype, comprising a determination of the *FokI* biallelic polymorphism, (wherein a ff genotype is designated as the presence of the restriction site on both alleles and a FF genotype designates the absence of the restriction site on both alleles) or a marker in linkage disequelibrium therewith, and a determination of the genotype of another marker, wherein the combination of alleles at the *FokI* polymorphism and at the other marker, shows a significant association with breast cancer.

Another aim of the present invention is to use a combination of polymorphisms of the vitamin D receptor (VDR) gene or an equivalent thereof, and a second polymorphism of the VDR or an equivalent thereof, as markers for breast cancer susceptibility and/or response to breast cancer preventive or curative therapy. One combination of a polymorphism of the vitamin D receptor (VDR) gene, or any polymorphism in linkage disequilibrium therewith, when combined with another polymorphism of the VDR gene, or any polymorphism in linkage disequilibrium therewith, can be used as a test for breast cancer susceptibility or protection, for responsiveness to treatment of breast cancer, for breast cancer prognosis or severity, or as a means to classify patients in clinical trial for breast cancer (screening, diagnosis, prognosis or treatment).

In a particularly preferred embodiment, the VDR genotype is determined at two loci, the *BsmI* polymorphism (or at a locus in linkage disequilibrium therewith) and at an other locus of VDR In one especially preferred embodiment of the present invention this additional locus of VDR is the *FokI* polymorphism

Thus, in a particularly preferred embodiment of the present invention, the determination of the predisposition and/or protection and/or

prognosis of breast cancer is carried out by analyzing the combination of alleles at the *BsmI* and *FokI* polymorphisms of VDR, or combinations of alleles at loci in linkage disequilibrium with the *BsmI* and *FokI* polymorphisms.

In accordance with the present invention, such a determination of a combination of alleles at the *BsmI* and *FokI* polymorphisms enables the identification of a striking association with breast cancer. In a particularly preferred embodiment, the identification of the combination of allele BBff is strongly associated with breast cancer in a very signifcant fraction of breast cancer patients, while the BBFf and BBFF are shown to have a protecting effect against breast cancer.

The present invention indeed provides, in accordance with a preferred embodiment thereof, the strongest susceptibility marker for breast cancer ever identified, as it explains ten times more cases of the disease than the BRCA1 and BRCA2 genes combined. Indeed, when the number of total cases of breast cancer attributable to these two VDR polymorphisms was computed, it was shown that 45% of all cases of breast cancer were attributable to VDR gene polymorphism, as detected by the combined testing of the *BsmI* and the *FokI* genotypes.

The present invention also relates to vectors, including expression vectors harboring a VDR gene (or fragment or fusion thereof) having a genotype in accordance with the present invention (i.e. a predisposing genotype, Bbff, or alternatively, a protecting genotype, BBFF; or other genotypes isolated from patients or genetically engineered), cells harboring such vectors, and non-human animals harboring such vectors or cells.

One of a polymorphism of the VDR gene, or any polymorphism in linkage disequilibrium therewith, when combined with another polymorphism of the VDR gene, or any polymorphism in linkage disequilibrium therewith, can further be used as a test for screening drugs for breast cancer or for determining the best treatment therefor.

Broadly, the present invention aims at providing a method of determining at least two polymorphisms in the vitamin D receptor gene, wherein this determination can be correlated with a predisposition or a protection to breast cancer. This determination can be based on a variety of genotyping methods at the DNA, RNA or protein level.

Another aim of the present invention is to provide a method of prognosing and/or forecasting the development of breast cancer in a patient, which comprises determining at least two polymorphisms of the VDR gene of the present invention, associated with breast cancer, or any polymorphism in linkage disequilibrium therewith, in a biological sample of the patient, wherein a combination of the polymorphisms at the VDR loci shows a significant association with breast cancer.

Another aim of the present invention is to provide means of identifying young women that will be at risk of developing breast cancer and to categorize those that are likely to respond significantly to preventive therapy. An aim of the present invention is thus to provide means of identification of target sub-groups of women for breast cancer prevention measures/programs.

Another aim of the present invention is to provide means to determine which sub-group of women will most benefit from breast cancer treatment(s) and eventually predict their response to therapy or choose the optimal preventive pharmacotherapy.

Another aim of the present invention is to identify means of predicting and managing interventions for breast cancer as well as identifying and/or characterizing biological parameters which could enable the establishment of population-based breast cancer prevention and intervention programs.

In addition, it is an aim of the present invention to provide a method of selecting alleles of the VDR gene or in linkage disequilibrium therewith, which are suitable for designing an assay to screen compounds which can modulate the activity of a vitamin D receptor

Another aim of the present invention is to provide an assay to screen for drugs for the treatment and/or prevention of breast cancer. Having identified alleles which predispose to breast cancer (and those which predispose to a "resistance" to breast cancer), assays can be set-up to screen agents and select drugs which could be used in the treatment or prevention of breast cancer. Since some alleles of the VDR have been shown to affect the functionality of the vitamin D receptor (Tut et al. 1997, J. Clin. Endocrinol. 89(11):3777-3782), assays could be designed based on chosen genotypes of the VDR gene. A non-limiting example of a type of assay which could be designed includes, *cis-trans* assays similar to those described in USP 4,981,784. For example, a *cis-trans* assay could be set-up, based on the use of a chosen alleles of VDR, shown here to predispose to breast cancer (i.e. the combination of allele BBff of the VDR gene) as compared to a alleles of VDR, shown here to be associated with lower risk and/or protection of breast cancer (i.e. the combination of allele BBFF or BBFf of the VDR gene), and used to screen compounds. A non-limiting example of such an assay could be based on 2 cell lines (one expressing a predisposing genotype of VDR and one expressing a non-predisposing and/or protecting genotype of VDR) which could be used in parallel to screen for VDR-function modulating compounds. Of course, it will be understood that the cell line expressing the non-predisposing genotype of VDR (BBFF, for example) can be used as a positive control for the functionality of the vitamin D receptor.

It is thus an aim of the present invention to provide the means to identify compounds which could positively modulate the function of VDR having a breast cancer predisposing genotype (such as Bbff), to the level of the protecting genotype thereof (such as BBFF).

In a particular embodiment, such assays can be designed using cells from patients having a known genotype at the loci of the present invention, these cells harboring recombinant vectors could enable an

assessment of the functionality of the VDR and dissect the structure-function relationship of the vitamin D receptor and its role in breast cancer.

It shall be understood that the polymorphisms of the VDR and/or the determination of allelic variations in the VDR gene can be combined to the determination of allelic variations in other genes/markers linked to the predisposition to breast cancer and/or responsiveness to therapy therefor. This combination of genotype analyses could lead to better diagnoses programs and/or treatment of breast cancer. Non-limiting examples of such markers include BRCA1 and BRCA2

It shall also be understood that although breast cancer is significantly more preponderant in women, it can also be a deadly disease in men. Thus, the present invention is meant to also cover men.

In accordance with the present invention, there is therefore provided a method of determining an individual's predisposition to breast cancer, protection to breast cancer, development of breast cancer and/or responsiveness to therapy for breast cancer, which comprises determining a genotype of the vitamin D receptor gene (directly or indirectly by linkage disequilibrium) in a biological sample of the individual and analyzing allelic variation in the vitamin D receptor of the individual, thereby determining an individual's predisposition to breast cancer, protection to breast cancer, development of breast cancer and/or responsiveness to therapy therefor.

In accordance with the present invention there is provided a method for determining susceptibility to breast cancer, and/or response to therapy therefor The method comprises the step of determining the vitamin D receptor genotype of the individual, thereby determining an individual's susceptibility to breast cancer and/or response to therapy therefor.

Numerous methods for determining a genotype are known and availble to the skilled artisan. All these genotype determination methods are within the scope of the present invention.

Non-limiting examples of genotype determination include a restriction endonuclease digestion, a hybridization with allele specific oligonucleotides, a sequencing of the polymorphism, and an amplification of a segment of the vitamin D receptor (i.e. by PCR)

In accordance with the present invention, there is therefore provided a method of determining an individual's predisposition to breast cancer, protection to breast cancer, development of breast cancer and/or responsiveness to therapy therefor, which comprises determining vitamin D receptor polymorphism at at least one of the *BsmI* and *FokI* polymorphic loci (directly or indirectly using a marker in linkage disequilibrium with these polymorphic loci) in a biological sample of the individual and analyzing allelic variation in the vitamin D receptor gene of the individual, thereby determining an individual's predisposition to breast cancer, protection to breast cancer, development of breast cancer and/or responsiveness to therapy therefor.

In accordance with one embodiment of the invention, there is provided a specific model for use in prediction of breast cancer susceptibility and prognosis. The model comprises two vitamin D receptor gene polymorphisms at the *BsmI* and *FokI* loci thereof, that allow to identify a subset of women that are at significantly increased risk of breast cancer as compared to those bearing other variants of this gene

In accordance with a preferred embodiment of the present invention, a single gene, the vitamin D receptor gene, has been identified as such a target to assess this predisposition.

In accordance with the present invention, the vitamin D receptor polymorphism, without limitation, is selected from the *BsmI* polymorphism located in the last intron of the VDR gene and the *FokI* polymorphism located in the first exon of the s gene, or any DNA variant or mutation which shows some degree of linkage disequelibrium with one of these polymorphisms.

In some embodiments of the present invention, at least one pair of primers is designed to specifically amplify a segment of the vitamin D receptor. In another embodiment, the region of the VDR gene which is amplified is in exon 1 and/or in in the last intron thereof This at least one pair of primers is preferably derived from a nucleic acid sequence of the vitamin D receptor gene or flanking portion thereof, to amplify a segment of the vitamin D receptor gene, as commonly known. Of course, other primer pairs can be designed, based on the known sequence of the VDR gene Method to design primer pairs form known sequences are commonly known in the art.

In accordance with a preferred embodiment of the present invention, primers used for amplifying a segment of the vitamin D receptor are defined as:

5'-CAACCAAGAC TACAAGTACC GCGTCAGTGA-3' (SEQ ID NO:1) and 5'-TATCGTGAGT AAGGCAGGAG AGGGAGACC-3' (SEQ ID NO:2); or

5'-AGCTGGCCCT GGCACTGACT CTGCTCT-3' (SEQ ID NO:3) and 5'-ATGGAAACAC CTTGCTTCTT CTCCCTC-3' (SEQ ID NO:4).

In accordance with a preferred embodiment, the method of the present invention includes detecting the vitamin D receptor polymorphisms by analyzing the restriction fragment length polymorphisms using an endonuclease digestion. The method can further include a step prior to the vitamin D receptor gene digestion, wherein at least a fragment of the vitamin D receptor is amplified, for example, by polymerase chain reaction. The step of determining the vitamin D receptor genotype could also comprise hybridizing with allele specific oligonucleotides.

Suitable endonucleases for genotyping the vitamin D receptor gene, in accordance with a preferred embodiment of the present invention are known in the art. Non-limiting examples thereof include, *BsmI*, *ApaI*, *TaqI*, *FokI* and their isoschizomers.

In accordance with another embodiment of the present invention, the vitamin D receptor genotype can be determined using a

polymorphic variant site in linkage disequilibrium with at least one allelic variant as detected with *BsmI*, *ApaI*, *TaqI*, *FokI*, and isoschizomers thereof, in the restriction endonuclease digestion, or otherwise.

The polymorphism of the vitamin D receptor gene can be detected using at least one oligonucleotide specific to the normal or variant vitamin D receptor gene allele. Methods to design specific probes from a known nucleic acid sequence are commonly known in the art.

The present invention also provides a kit for determining predisposition to low, intermediate or high risk of breast cancer or to a protection to breast cancer of a patient, which includes at least a probe specific for the vitamin D receptor; a polymorphism selected from: a) a *BsmI* polymorphism; b) a FokI polymorphism; c) a polymorphism of VDR showing a significant association with breast cancer; and d) a polymorphism in linkage disequilibrium with the polymorphisms of a)-c).

In one embodiment, the present invention provides a specific detection of a VDR polymorphism of the VDR gene using a nucleic acid for the specific detection of this VDR polymorphism in a sample comprising at a nucleic acid sequence which binds under stringent conditions to the VDR polymorphic nucleic acid sequence.

In one prefered embodiment, the present invention relates to nucleic acid probes which are complementary to a VDR polymorphic sequence, consisting of at least 10 consecutive nucleotides (preferably, 15, 20, 25, or 30) and which specifically hybridize to the VDR polymorphic sequence under high stringency condition. A non-limiting example of a polymorphic specific probe according to the invention includes a probe which would bind to the *FokI* polymorphic sequence (specific to the C to T transition, creating a new initiator AUG giving rise to a protein having a three amino acids difference).

In one embodiment of the above described method, a nucleic acid probe is immobilized on a solid support Non-limiting examples of solid supports include plastics (i.e.polycarbonate), acrylic resins (i.e. polyacrylamide

and latex beads); and carbohydrates (i.e. agarose and sepharose). Techniques for coupling nucleic acid probes to solid supports are well known in the art.

Similarly to the probes of the present invention, the antibodies (i.e. an antibody specific to a polymorphism of VDR) of the present invention can be immobilized on a solid support. As known in the art, similar supports as those used for probe immobilization can be used for antibody immobilization on a solid support. Also well known in the art are the techniques for coupling antibodies to such solid supports. The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as in immunochromatography according to known methods.

Non-limiting examples of test samples suitable for carrying the methods of the present invention include, cells or nucleic acid extracts of cells, or biological fluids. Of course, the type of test sample used can vary according to the assay format, the method of detection, and the particular needs of the clinical practioner which will readily adapt the methods of preparation of the sample and the method of detection so that they are compatible, in accordance with the knowledge in the art

In accordance with one embodiment of the present invention, the allelic variation in the vitamin D receptor gene can be analyzed indirectly using a nucleic acid variant, or equivalent in linkage disequilibrium with one of the polymorphic sites of VDR The allelic variation in the vitamin D receptor gene can also be analyzed directly by determining the genotype within the vitamin D receptor gene

In accordance with the present invention, the polymorphism of the vitamin D receptor gene can be used as a marker for breast cancer susceptibility. The polymorphism in linkage disequilibrium with the markers used can also be used as a test for breast cancer susceptibility, breast cancer protection, or for responsiveness to treatment for breast cancer, for breast cancer prognosis or severity, or as a means to classify patients in clinical trials for breast cancer (screening, diagnosis, prognosis or treatment).

In order to provide a clear and consistent understanding of terms used in the present description, a number of definitions are provided hereinbelow

As used herein the term "RFLP" refers to restriction fragment length polymorphism.

The terms "polymorphism", "DNA polymorphism" and the like, refer to any sequence in the human genome which exists in more than one version or variant in the population.

The term "linkage disequilibrium" refers to any degree of non-random genetic association between one or more allele(s) of two different polymorphic DNA sequences, that is due to the physical proximity of the two loci. Linkage disequilibrium is present when two DNA segments that are very close to each other on a given chromosome will tend to remain unseparated for several generations with the consequence that alleles of a DNA polymorphism (or marker) in one segment will show a non-random association with the alleles of a different DNA polymorphism (or marker) located in the other DNA segment nearby. Hence, testing of one of a marker in linkage desiquilibrium with the polymorphisms of the present invention at the VDR gene (indirect testing), will give almost the same information as testing for the herein-identified polymorphisms of the VDR gene directly This situation is encountered throughout all the human genome when two DNA polymorphisms that are very close to each other are studied. Such a linkage disequilibrium has been reported with several polymorphisms in several genes (i.e the vitamin D receptor gene [Morrisson et al., 1994, Nature 367:284-287]). Various degrees of linkage disequilibrium can be encountered between two genetic markers so that some are more closely associated than others.

The terms "vitamin D receptor polymorphism" or "genetic marker" are intended to include, without limitation, *BsmI*, *TaqI*, *ApaI* or *FokI* and any other allelic variant of the vitamin D receptor gene that shows some degree

of linkage disequilibrium in any population sub-group with at least one of the above-mentioned vitamin D receptor polymorphisms.

The vitamin D receptor gene polymorphism sites in accordance with the present invention can be located within the vitamin D receptor gene, or on each side thereof, provided that is on the same chromosome and in linkage disequilibrium with the VDR polymorphism of the present invention. Distances between markers in linkage disequilibrium can vary widely (below 50 kb to more than 1 mega base) depending on the genetic structure of the population and is ascertainable by a statistically significant association between the markers.

It shall be recognized by the person skilled in the art to which the present invention pertains, that since some of the polymorphisms herein identified in the VDR gene can be within the coding region of the gene and therefore expressed (the *FokI* polymorphism, for example), that the present invention should not be limited to the identification of polymorphisms at the DNA level (whether on genomic DNA, amplified DNA, cDNA or the like). Indeed, the herein-identified polymorphisms could be detected at the mRNA or protein level. Such detections of polymorphism identification on mRNA or protein are known in the art. Non-limiting examples include detection based on oligos designed to hybridize to mRNA or ligands such as antibodies which are specific to the encoded polymorphism (i.e. specific to the 3 extra amino acids encoded by the *FokI* polymorphism for example).

Since some of the polymorphisms of the present invention are expressed, one of the advantages of the present invention is to enable a determination of the polymorphisms in the VDR gene, in easily obtainable cells which express these genes. A non-limiting example thereof is lymphocytes, thereby enabling a genotyping from a simple blood sample.

Nucleotide sequences are presented herein by single strand, in the 5' to 3' direction, from left to right, using the one letter nucleotide

symbols as commonly used in the art and in accordance with the recommendations of the IUPAC-IUB Biochemical Nomenclature Commission.

Unless defined otherwise, the scientific and technological terms and nomenclature used herein have the same meaning as commonly understood by a person of ordinary skill to which this invention pertains. Generally, the procedures for cell cultures, infection, molecular biology methods and the like are common methods used in the art. Such standard techniques can be found in reference manuals such as for example Sambrook et al. (1989, Molecular Cloning -A Laboratory Manual, Cold Spring Harbor Laboratories) and Ausubel et al. (1994, Current Protocols in Molecular Biology, Wiley, New York).

The present description refers to a number of routinely used recombinant DNA (rDNA) technology terms. Nevertheless, definitions of selected examples of such rDNA terms are provided for clarity and consistency.

As used herein, "nucleic acid molecule", refers to a polymer of nucleotides. Non-limiting examples thereof include DNA (i.e. genomic DNA, cDNA) and RNA molecules (i.e. mRNA) The nucleic acid molecule can be obtained by cloning techniques or synthesized. DNA can be double-stranded or single-stranded (coding strand or non-coding strand [antisense]).

The term "recombinant DNA" as known in the art refers to a DNA molecule resulting from the joining of DNA segments. This is often referred to as genetic engineering.

The term "DNA segment", is used herein, to refer to a DNA molecule comprising a linear stretch or sequence of nucleotides. This sequence when read in accordance with the genetic code, can encode a linear stretch or sequence of amino acids which can be referred to as a polypeptide, protein, protein fragment and the like.

The terminology "amplification pair" refers herein to a pair of oligonucleotides (oligos) of the present invention, which are selected to be used together in amplifying a selected nucleic acid sequence by one of a number of types of amplification processes, preferably a polymerase chain

reaction. Other types of amplification processes include ligase chain reaction, strand displacement amplification, or nucleic acid sequence-based amplification, as explained in greater detail below. As commonly known in the art, the oligos are designed to bind to a complementary sequence under selected conditions.

The nucleic acid (i.e. DNA or RNA) for practicing the present invention may be obtained according to well known methods.

Oligonucleotide probes or primers of the present invention may be of any suitable length, depending on the particular assay format and the particular needs and targeted genomes employed. In general, the oligonucleotide probes or primers are at least 12 nucleotides in length, preferably between 15 and 24 molecules, and they may be adapted to be especially suited to a chosen nucleic acid amplification system. As commonly known in the art, the oligonucleotide probes and primers can be designed by taking into consideration the melting point of hydrizidation thereof with its targeted sequence (see below and in Sambrook et al., 1989, Molecular Cloning -A Laboratory Manual, 2nd Edition, CSH Laboratories; Ausubel et al., 1989, in Current Protocols in Molecular Biology, John Wiley & Sons Inc., N.Y.).

The term "oligonucleotide" or "DNA" molecule or sequence refers to a molecule comprised of the deoxyribonucleotides adenine (A), guanine (G), thymine (T) and/or cytosine (C), in a double-stranded form, and comprises or includes a "regulatory element" according to the present invention, as the term is defined herein. The term "oligonucleotide" or "DNA" can be found in linear DNA molecules or fragments, viruses, plasmids, vectors, chromosomes or synthetically derived DNA. As used herein, particular double-stranded DNA sequences may be described according to the normal convention of giving only the sequence in the 5' to 3' direction.

"Nucleic acid hybridization" refers generally to the hybridization of two single-stranded nucleic acid molecules having complementary base sequences, which under appropriate conditions will form a thermodynamically favored double-stranded structure. Examples of

hybridization conditions can be found in the two laboratory manuals referred above (Sambrook et al., 1989, *supra* and Ausubel et al , 1989, *supra*) and are commonly known in the art. In the case of a hybridization to a nitrocellulose filter, as for example in the well known Southern blotting procedure, a nitrocellulose filter can be incubated overnight at 65°C with a labeled probe in a solution containing 50% formamide, high salt (5 x SSC or 5 x SSPE), 5 x Denhardt's solution, 1% SDS, and 100 µg/ml denatured carrier DNA (i.e. salmon sperm DNA). The non-specifically binding probe can then be washed off the filter by several washes in 0.2 x SSC/0 1% SDS at a temperature which is selected in view of the desired stringency: room temperature (low stringency), 42°C (moderate stringency) or 65°C (high stringency). The selected temperature is based on the melting temperature (Tm) of the DNA hybrid. Of course, RNA-DNA hybrids can also be formed and detected. In such cases, the conditions of hybridization and washing can be adapted according to well known methods by the person of ordinary skill. Stringent conditions will be preferably used (Sambrook et al.,1989, *supra*).

Probes of the invention can be utilized with naturally occurring sugar-phosphate backbones as well as modified backbones including phosphorothioates, dithionates, alkyl phosphonates and α-nucleotides and the like. Modified sugar-phosphate backbones are generally taught by Miller, 1988, Ann. Reports Med. Chem. 23:295 and Moran et al., 1987, Nucleic acid molecule. Acids Res., 14:5019. Probes of the invention can be constructed of either ribonucleic acid (RNA) or deoxyribonucleic acid (DNA), and preferably of DNA.

The types of detection methods in which probes can be used include Southern blots (DNA detection), dot or slot blots (DNA, RNA), and Northern blots (RNA detection). Although less preferred, labeled proteins could also be used to detect a particular nucleic acid sequence to which it binds. More recently, PNAs have been described (Nielsen et al 1999, Current Opin. Biotechnol. 10:71-75). PNAs could also be used to detect the polymorphisms

of the present invention. Other detection methods include kits containing probes on a dipstick setup and the like.

Although the present invention is not specifically dependent on the use of a label for the detection of a particular nucleic acid sequence, such a label might be beneficial, by increasing the sensitivity of the detection. Furthermore, it enables automation Probes can be labeled according to numerous well known methods (Sambrook et al , 1989, supra). Non-limiting examples of labels include $^{3}H$, $^{14}C$, $^{32}P$, and $^{35}S$. Non-limiting examples of detectable markers include ligands, fluorophores, chemiluminescent agents, enzymes, and antibodies. Other detectable markers for use with probes, which can enable an increase in sensitivity of the method of the invention, include biotin and radionucleotides. It will become evident to the person of ordinary skill that the choice of a particular label dictates the manner in which it is bound to the probe.

As commonly known, radioactive nucleotides can be incorporated into probes of the invention by several methods. Non-limiting examples thereof include kinasing the 5' ends of the probes using gamma $^{32}P$ ATP and polynucleotide kinase, using the Klenow fragment of Pol I of *E. coli* in the presence of radioactive dNTP (i.e. uniformly labeled DNA probe using random oligonucleotide primers in low-melt gels), using the SP6/T7 system to transcribe a DNA segment in the presence of one or more radioactive NTP, and the like.

As used herein, "oligonucleotides" or "oligos" define a molecule having two or more nucleotides (ribo or deoxyribonucleotides). The size of the oligo will be dictated by the particular situation and ultimately on the particular use thereof and adapted accordingly by the person of ordinary skill. An oligonucleotide can be synthesized chemically or derived by cloning according to well known methods.

As used herein, a "primer" defines an oligonucleotide which is capable of annealing to a target sequence, thereby creating a double

stranded region which can serve as an initiation point for DNA synthesis under suitable conditions.

Amplification of a selected, or target, nucleic acid sequence may be carried out by a number of suitable methods. See generally Kwoh et al., 1990, Am. Biotechnol. Lab. 8:14-25 Numerous amplification techniques have been described and can be readily adapted to suit particular needs of a person of ordinary skill. Non-limiting examples of amplification techniques include polymerase chain reaction (PCR), ligase chain reaction (LCR), strand displacement amplification (SDA), transcription-based amplification, the Qβ replicase system and NASBA (Kwoh et al., 1989, Proc, Natl. Acad. Sci. USA 86, 1173-1177, Lizardi et al., 1988, BioTechnology 6:1197-1202; Malek et al., 1994, Methods Mol. Biol., 28:253-260; and Sambrook et al., 1989, *supra*). Preferably, amplification will be carried out using PCR.

Polymerase chain reaction (PCR) is carried out in accordance with known techniques. See, e.g., U.S. Pat. Nos. 4,683,195; 4,683,202; 4,800,159; and 4,965,188 (the disclosures of all three U.S. Patent are incorporated herein by reference). In general, PCR involves, a treatment of a nucleic acid sample (e.g., in the presence of a heat stable DNA polymerase) under hybridizing conditions, with one oligonucleotide primer for each strand of the specific sequence to be detected. An extension product of each primer which is synthesized is complementary to each of the two nucleic acid strands, with the primers sufficiently complementary to each strand of the specific sequence to hybridize therewith. The extension product synthesized from each primer can also serve as a template for further synthesis of extension products using the same primers. Following a sufficient number of rounds of synthesis of extension products, the sample is analysed to assess whether the sequence or sequences to be detected are present. Detection of the amplified sequence may be carried out by visualization following EtBr staining of the DNA following gel electrophores, or using a detectable label in accordance with known techniques,

and the like. For a review on PCR techniques (see PCR Protocols, A Guide to Methods and Amplifications, Michael et al. Eds, Acad. Press, 1990).

Ligase chain reaction (LCR) is carried out in accordance with known techniques (Weiss, 1991, Science 254:1292). Adaptation of the protocol to meet the desired needs can be carried out by a person of ordinary skill. Strand displacement amplification (SDA) is also carried out in accordance with known techniques or adaptations thereof to meet the particular needs (Walker et al., 1992, Proc. Natl. Acad. Sci. USA 89:392-396; and ibid., 1992, Nucleic Acids Res. 20:1691-1696).

As used herein, the term "gene" is well known in the art and relates to a nucleic acid sequence defining a single protein or polypeptide. A "structural gene" defines a DNA sequence which is transcribed into RNA and translated into a protein having a specific amino acid sequence thereby giving rise the a specific polypeptide or protein.

A "heterologous" (i.e. a heterologous gene) region of a DNA molecule is a subsegment segment of DNA within a larger segment that is not found in association therewith in nature. The term "heterologous" can be similarly used to define two polypeptidic segments not joined together in nature Non-limiting examples of heterologous genes include reporter genes such as luciferase, chloramphenicol acetyl transferase, β-galactosidase, and the like which can be juxtaposed or joined to heterologous control regions or to heterologous polypeptides.

The term "vector" is commonly known in the art and defines a plasmid DNA, phage DNA, viral DNA and the like, which can serve as a DNA vehicle into which DNA of the present invention can be cloned. Numerous types of vectors exist and are well known in the art.

The term "expression" defines the process by which a gene is transcribed into mRNA (transcription), the mRNA is then being translated (translation) into one polypeptide (or protein) or more.

The terminology "expression vector" defines a vector or vehicle as described above but designed to enable the expression of an inserted sequence following transformation into a host The cloned gene (inserted sequence) is usually placed under the control of control element sequences such as promoter sequences. The placing of a cloned gene under such control sequences is often refered to as being operably linked to control elements or sequences.

Operably linked sequences may also include two segments that are transcribed onto the same RNA transcript. Thus, two sequences, such as a promoter and a "reporter sequence" are operably linked if transcription commencing in the promoter will produce an RNA transcript of the reporter sequence. In order to be "operably linked" it is not necessary that two sequences be immediately adjacent to one another.

Expression control sequences will vary depending on whether the vector is designed to express the operably linked gene in a prokaryotic or eukaryotic host or both (shuttle vectors) and can additionally contain transcriptional elements such as enhancer elements, termination sequences, tissue-specificity elements, and/or translational initiation and termination sites.

Prokaryotic expressions are useful for the preparation of large quantities of the protein encoded by the DNA sequence of interest. This protein can be purified according to standard protocols that take advantage of the intrinsic properties thereof, such as size and charge (i.e. SDS gel electrophoresis, gel filtration, centrifugation, ion exchange chromatography...). In addition, the protein of interest can be purified via affinity chromatography using polyclonal or monoclonal antibodies The purified protein can be used for therapeutic applications.

The DNA construct can be a vector comprising a promoter that is operably linked to an oligonucleotide sequence of the present invention, which is in turn, operably linked to a heterologous gene, such as the gene for the

luciferase reporter molecule. "Promoter" refers to a DNA regulatory region capable of binding directly or indirectly to RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of the present invention, the promoter is bound at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter will be found a transcription initiation site (conveniently defined by mapping with S1 nuclease), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boses and "CCAT" boxes. Prokaryotic promoters contain Shine-Dalgarno sequences in addition to the -10 and -35 consensus sequences.

In accordance with one embodiment of the present invention, an expression vector can be constructed to assess the functionality of specific alleles of the VDR gene and of the interaction of such alleles. Non-limiting examples of such expression vectors include a vector comprising the vitamin D responsive element (the cis sequences [i.e. DNA sequence to which a factor binds] enabling vitamin D-dependent modulating effects of promoter activity are known in the art) operably linked to a chosen promoter and modulating the activity thereof, the promoter driving the expression of a reporter gene. When such a vector is tranfected in a cell expressing VDR, the modulating effect of the promoter activity can be assessed by determining the level of expression of the reporter gene. In one embodiment, the vector is transfected into a cell of a patient having the genotype of VDR shown herein to be associated with a low risk and/or protection of breast cancer, or in a cell from a patient having the genotype of VDR shown herein to be associated with a moderate or high risk of breast cancer. These cells can serve to screen for compounds that modulate the promoter activity, in order to identify compounds that could be used to treat patients predicted to be at moderate or high risk of breast cancer Of course, it will be understood that the VDR gene expressed

by these cells can be modified at will (i.e. by *in vitro* mutagenesis or the like). Similarly, numerous combinations of genotypes can be tested in such assays to dissect the functional relationship between the VDR genotype and its function in vitamin D-dependent function and/or its function in breast cancer. It will also be clear to the skilled artisan, that such indicator cells expressing VDR, could also be engineered by choosing a cell line and transfecting thereinto, chosen genotypes of VDR and one expression vector as described above. Non-human transgenic animals expressing chosen alleles of VDR could also be prepared and used to screen compounds that affect vitamin D receptor function and possibly overcome a predisposition to breast cancer, perhaps to the level observed with the BBFF genotype.

As used herein, the designation "functional derivative" denotes, in the context of a functional derivative of a sequence whether an nucleic acid or amino acid sequence, a molecule that retains a biological activity (either function or structural) that is substantially similar to that of the original sequence. This functional derivative or equivalent may be a natural derivative or may be prepared synthetically. Such derivatives include amino acid sequences having substitutions, deletions, or additions of one or more amino acids, provided that the biological activity of the protein is conserved. The same applies to derivatives of nucleic acid sequences which can have substitutions, deletions, or additions of one or more nucleotides, provided that the biological activity of the sequence is generally maintained. When relating to a protein sequence, the substituting amino acid as chemico-physical properties which are similar to that of the substituted amino acid The similar chemico-physical properties include, similarities in charge, bulkiness, hydrophobicity, hydrophylicity and the like The term "functional derivatives" is intended to include "fragments", "segments", "variants", "analogs" or "chemical derivatives" of the subject matter of the present invention.

Thus, the term "variant" refers herein to a protein or nucleic acid molecule which is substantially similar in structure and biological activity to the protein or nucleic acid of the present invention.

The functional derivatives of the present invention can be synthesized chemically or produced through recombinant DNA technology. all these methods are well known in the art.

As used herein, "chemical derivatives" is meant to cover additional chemical moieties not normally part of the subject matter of the invention. Such moieties could affect the physico-chemical characteristic of the derivative (i.e. solubility, absorption, half life and the like, decrease of toxicity). Such moieties are examplified in Remington's Pharmaceutical Sciences (1980). Methods of coupling these chemical-physical moieties to a polypeptide are well known in the art.

The term "allele" defines an alternative form of a gene which occupies a given locus on a chromosome.

As commonly known, a "mutation" is a detectable change in the genetic material which can be transmitted to a daughter cell. As well known, a mutation can be, for example, a detectable change in one or more deoxyribonucleotide. For example, nucleotides can be added, deleted, substituted for, inverted, or transposed to a new position. Spontaneous mutations and experimentally induced mutations exist. The result of a mutations of nucleic acid molecule is a mutant nucleic acid molecule. A mutant polypeptide can be encoded from this mutant nucleic acid molecule.

As used herein, the term "purified" refers to a molecule having been separated from a cellular component. Thus, for example, a "purified protein" has been purified to a level not found in nature. A "substantially pure" molecule is a molecule that is lacking in all other cellular components.

As used herein, the terms "molecule", "compound", or "agent" are used interchangeably and broadly to refer to natural, synthetic or semi-synthetic molecules or compounds. The term "molecule" therefore

denotes for example chemicals, macromolecules, cell or tissue extracts (from plants or animals) and the like. Non limiting examples of molecules include nucleic acid molecules, peptides, ligands, including antibodies, carbohydrates and pharmaceutical agents. The agents can be selected and screened by a variety of means including random screening, rational selection and by rational design using for example protein or ligand modelling methods such as computer modelling. The terms "rationally selected" or "rationally designed" are meant to define compounds which have been chosen based on the configuration of the interaction domains of the present invention. As will be understood by the person of ordinary skill, macromolecules having non-naturally occurring modifications are also within the scope of the term "molecule". For example, peptidomimetics, well known in the pharmaceutical industry and generally referred to as peptide analogs can be generated by modelling as mentioned above. Similarly, in a preferred embodiment, the polypeptides of the present invention are modified to enhance their stability. It should be understood that in most cases this modification should not alter the biological activity of the protein. The molecules identified in accordance with the teachings of the present invention have a therapeutic value in diseases or conditions in which a apparently lower activity and/or level of the VDR is linked to a genotype of VDR identified in accordance with the present invention Alternatively, the molecules identified in accordance with the teachings of the present invention find utility in the development of compounds which can modulate the activity and/or level of the vitamin D receptor in an animal and/or overcome a predisposition to breast cancer.

As used herein, agonists and antagonists also include potentiators of known compounds with such agonist or antagonist properties. In one embodiment, modulators of the level or the activity of the VDR can be identified and selected by contacting the indicator cell with a compound or mixture or library of molecules for a fixed period of time. In certain

embodiments, the "breast cancer-low risk-associated alleles"of the VDR gene can be used as positive controls.

An indicator cell in accordance with the present invention can be used to identify antagonists. For example, the test molecule or molecules are incubated with the host cell in conjunction with one or more agonists held at a fixed concentration. An indication and relative strength of the antagonistic properties of the molecule(s) can be provided by comparing the level of gene expression in the indicator cell in the presence of the agonist, in the absence of test molecules vs in the presence thereof. Of course, the antagonistic effect of a molecule can also be determined in the absence of agonist, simply by comparing the level of expression of the reporter gene product in the presence and absence of the test molecule(s).

It shall be understood that the *"in vivo"* experimental model can also be used to carry out an *"in vitro"* assay. For example, cellular extracts from the indicator cells can be prepared and used in an *"in vitro"* test. A non-limiting example thereof include binding assays.

As used herein the recitation "indicator cells" refers to cells that express a given genotype of VDR according to the present invention. As alluded to above, such indicator cells can be used in the screening assays of the present invention. In certain embodiments, the indicator cells have been engineered so as to express a chosen derivative, fragment, homolog, or mutant of a genotype of the present invention. The cells can be yeast cells or higher eukaryotic cells such as mammalian cells. In one particular embodiment, the indicator cell would be a yeast cell harboring vectors enabling the use of the two hybrid system technology, as well known in the art (Ausubel et al., 1994, *supra*) and can be used to test a compound or a library thereof In another embodiment, the *cis-trans* assay as described in USP 4,981,784, can be adapted and used in accordance with the present invention. Such an indicator cell could be used to rapidly screen at high-throughput a vast array of test molecules. In a particular embodiment, the reporter gene is luciferase or β-Gal.

In some embodiments, it might be beneficial to express a fusion protein. The design of constructs therefor and the expression and production of fusion proteins and are well known in the art (Sambrook et al., 1989, *supra*; and Ausubel et al., 1994, *supra*).

Non limiting examples of such fusion proteins include a hemaglutinin fusions and Gluthione-S-transferase (GST) fusions and Maltose binding protein (MBP) fusions. In certain embodiments, it might be beneficial to introduce a protease cleavage site between the two polypeptide sequences which have been fused. Such protease cleavage sites between two heterologously fused polypeptides are well known in the art.

In certain embodiments, it might also be beneficial to fuse the protein of the present invention to signal peptide sequences enabling a secretion of the fusion protein from the host cell Signal peptides from diverse organisms are well known in the art. Bacterial OmpA and yeast Suc2 are two non limiting examples of proteins containing signal sequences. In certain embodiments, it might also be beneficial to introduce a linker (commonly known) between the interaction domain and the heterologous polypeptide portion. Such fusion protein find utility in the assays of the present invention as well as for purification purposes, detection purposes and the like.

For certainty, the sequences and polypeptides useful to practice the invention include without being limited thereto mutants, homologs, subtypes, alleles and the like. It shall be understood that generally, the sequences of the present invention should encode a functional (albeit defective) VDR. It will be clear to the person of ordinary skill that whether the VDR sequence of the present invention, variant, derivative, or fragment thereof retains its function, can be determined by using the teachings and assays of the present invention and the general teachings of the art

It should be understood that the VDR protein of the present invention can be modified, for example by *in vitro* mutagenesis, to dissect the structure-function relationship thereof and permit a better design and

identification of modulating compounds. However, some derivative or analogs having lost their biological function may still find utility, for example for raising antibodies These antibodies could be used for detection or purification purposes. In addition, these antibodies could also act as competitive or non-competitive inhibitor and be found to be modulators of the activity of the VDR protein of the present invention.

A host cell or indicator cell has been "transfected" by exogenous or heterologous DNA (e.g a DNA construct) when such DNA has been introduced inside the cell. The transfecting DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transfecting DNA may be maintained on a episomal element such as a plasmid. With respect to eukaryotic cells, a stably transfected cell is one in which the transfecting DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transfecting DNA. Transfection methods are well known in the art (Sambrook et al., 1989, *supra*; Ausubel et al., 1994 *supra*). The use of a mammalian cell as indicator can provide the advantage of furnishing an intermediate factor, which permits for example the interaction of two polypeptides which are tested, that might not be present in lower eukaryotes or prokaryotes. It will be understood that extracts from mammalian cells for example could be used in certain embodiments, to compensate for the lack of certain factors.

In general, techniques for preparing antibodies (including monoclonal antibodies and hybridomas) and for detecting antigens using antibodies are well known in the art (Campbell, 1984, In "Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology", Elsevier Science Publisher, Amsterdam, The Netherlands) and in Harlow et al., 1988 (in. Antibody-A Laboratory Manual, CSH Laboratories). The present

invention also provides polyclonal, monoclonal antibodies, or humanized versions thereof, chimeric antibodies and the like which inhibit or neutralize their respective interaction domains and/or are specific thereto.

From the specification and appended claims, the term therapeutic agent should be taken in a broad sense so as to also include a combination of at least two such therapeutic agents. Further, the DNA segments or proteins according to the present invention could be introduced into individuals in a number of ways. For example, cells can be isolated from the afflicted individual, transformed with a DNA construct according to the invention and reintroduced to the afflicted individual in a number of ways. Alternatively, the DNA construct can be administered directly to the afflicted individual. The DNA construct can also be delivered through a vehicle such as a liposome, which can be designed to be targeted to a specific cell type, and engineered to be administered through different routes. For example, a vitamin D receptor gene having the genotype associated with low risk of breast cancer could be introduced in cells or in an individual displaying the VDR polymorphism associated with high risk of breast cancer

For administration to humans, the prescribing medical professional will ultimately determine the appropriate form and dosage for a given patient, and this can be expected to vary according to the chosen therapeutic regimen (i.e DNA construct, protein, cells), the response and condition of the patient as well as the severity of the disease.

Composition within the scope of the present invention should contain the active agent (i.e. molecule, hormone) in an amount effective to achieve the desired therapeutic effect while avoiding adverse side effects. Typically, the nucleic acids in accordance with the present invention can be administered to mammals (i.e humans) in doses ranging from 0.005 to 1 mg per kg of body weight per day of the mammal which is treated. Pharmaceutically acceptable preparations and salts of the active agent are within the scope of the present invention and are well known in the art (Remington's Pharmaceutical

Science, 16th Ed , Mack Ed.). For the administration of polypeptides, antagonists, agonists and the like, the amount administered should be chosen so as to avoid adverse side effects. The dosage will be adapted by the clinician in accordance with conventional factors such as the extent of the disease and different parameters from the patient Typically, 0 001 to 50 mg/kg/day will be administered to the mammal.

The present invention relates to a kit for assessing a predisposition to breast cancer comprising a determination of the genotype at the VDR locus (or a locus in linkage desiquilibrium therewith) using a nucleic acid fragment, a protein or a ligand, or a restriction enzyme in accordance with the present invention For example, a compartmentalized kit in accordance with the present invention includes any kit in which reagents are contained in separate containers Such containers include small glass containers, plastic containers or strips of plastic or paper Such containers allow the efficient transfer of reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another Such containers will include in one particular embodiment a container which will accept the test sample (DNA protein or cells), a container which contains the primers used in the assay, containers which contain enzymes, containers which contain wash reagents, and containers which contain the reagents used to detect the extension products.

It will be readily recognized by the person of ordinary skill, that the nucleic acid sequences, probes, primers, antibodies and the like of the present invention enabling a detection of a VDR polymorphism of the present invention can be incorporated into anyone of numerous established kit formats which are well known in the art.

Other objects, advantages and features of the present invention will become more apparent upon reading of the following

non-restrictive description of preferred embodiments which is exemplary and should not be interpreted as limiting the scope of the present invention.

## DESCRIPTION OF THE PREFERRED EMBODIMENT

In accordance with a preferred embodiment of the present invention, there is provided a specific model for use in prediction of breast cancer susceptibility and prognosis and consists in vitamin D receptor (VDR) gene polymorphisms that allow, individually (for one of the two) or in combination, to identify a subset of women that is at significantly increased risk of breast cancer as compared to those bearing other variants of this gene.

In accordance with a preferred embodiment of the present invention, a single gene, the vitamin D receptor (VDR) gene has been identified. The polymorphism of this gene is associated with a significant proportion of breast cancer cases in the general population (up to 40% of all cases (attributable risk)). Polymorphisms of this gene, namely the *BsmI* polymorphism located in the last intron of the VDR gene and the *FokI* polymorphism located in the first exon, are tested according to published protocols as described hereinafter

**DNA purification**

Blood samples were drawn into Vacutaner™ containing EDTA and 200 µl was aliquoted into 1.5 ml Eppendorf™ tubes within 48 hours and stored at -20°C until DNA purification Genomic DNA was isolated from peripheral blood leukocytes by a mini-method necessitating only 200 µl of whole blood where all steps are processed in a single 1.5 ml tube (Rousseau et al. *Hum. Mut.*, 4:51-54, 1994). Isolated DNA (5-7 µg) was resuspended into 100 µl TE 20:5 buffer (20mM Tris, 5mM EDTA), heated at 65°C for 4 hours and stored at 4°C until PCR was performed

**VDR genotype analysis**

VDR genotype was first assessed for *BsmI*, *ApaI* and *TaqI* polymorphisms as described by Morisson et al. (Morrison et al., *Nature*,

367:284-287, 1994 and USP 5,593,033) after amplification of a single 2.3 kb fragment by polymerase chain reaction (PCR) spanning 3' end of exon 8 to 5' end of non-translated exon 9, using forward primer (P1) 5'-CAACCAAGAC TACAAGTACC GCGTCAGTGA-3' (SEQ ID NO:1)(Morrison et al., *Nature*, 367:284-287, 1994) and reverse primer (P2) 5'-TATCGTGAGT AAGGCAGGAG AGGGAGACC-3' (SEQ ID NO.2). PCR was carried out in a Perkin-Elmer 480 DNA thermal cycler™ (Perkin-Elmer Corporation, Norwalk, CT). Genomic DNA (200 ng) was amplified through 35 cycles in 50 µl containing 1 µM of the two primers, 200 µM of the four deoxyribonucleotides and 2.5 U of Taq polymerase and its buffer (Promega Corporation, Madison, WI) in 1.2 mM $MgCl_2$. PCR conditions included an initial denaturation of 7 min. at 96°C followed by 35 cycles of amplification with denaturation at 94°C for 60 sec., annealing at 60°C for 60 sec. and extension at 72°C for 4 min. A final extension step at 72°C for 10 min. was added.

About 400 ng of the 2.3 Kb PCR product were digested overnight in three different test tubes for three polymorphic sites at 65°C (*BsmI*), 25°C (*ApaI*) and 37°C (*TaqI*)(New England Biolabs, Beverly, Ma) for 40 subjects. Following endonuclease digestion, genotypes were determined by ethidium bromide staining and UVB transillumination after a 2 % agarose gel electrophoresis. The presence of the three polymorphisms and the strong linkage disequilibrium already described (Morrison et al., *Nature*, 367:284-287, 1994) with 36 of the 40 subjects for which the anticipated haplotypes could be inferred or were compatible with the observed combinations of genotypes (BAt or baT, where b,a,t represent the presence of polymorphic sites and B,A,T the absence of the polymorphic sites).

Subsequently, endonuclease digestion of amplified genomic DNA was performed only with *BsmI* using the same exon 8 forward primer but an intron VIII reverse primer (P3) 5'-AACCAGCGGG AAGAGGTCAA GGG-3' (SEQ ID NO:3) instead (Morrison et al , *Nature*, 367:284-287, 1994). PCR conditions were basically the same except that $MgCl_2$ concentration was

2.0 mM and polymerase extension was decreased to only two minutes. After *BsmI* digestion, genotypes were again visualized by ethidium bromide after migration in a 2 % agarose gel electrophoresis. Absence of the polymorphic site (B) resulted in a 850 bp fragment while presence of the polymorphic site (b) resulted in 700 bp and 150 bp fragments (Morrison et al., *Nature*, 367:284-287, 1994).

A C-to-T transition polymorphism in the VDR resulting in an initiation codon (ATG) three codons proximal to a downstream start site and causing a three amino acids difference in the VDR sequence was previously described (Saijo et al., *Am J. Hum Genet.* 48:668-673, 1991). This polymorphism creates a *FokI* site and was also tested for 543 subjects. PCR and digestion was conducted as described by Gross et al. (Gross et al., *J. Bone Miner. Res.* 11.1850-1855, 1996) except that digestion was achieved with 1 U of *FokI* and was allowed overnight to minimize partial digests.

All genotype readings were performed by three independent readers blind to the disease status of the participants and only genotypes for which all three readers were concordant were kept for analysis, the others being declared conflictual. Perfect concordance between readers was over 97% of all samples.

It was thus discovered in accordance with the invention that testing for these polymorphisms in the VDR receptor gene allows to distinguish between women at lower risk of breast cancer and those at higher risk of the disease. According to the combined *BsmI* and *FokI* genotypes as determined by genetic testing of those two genetic markers, the odds ratio of a women of developing a breast cancer varies between 1 (for the lowest risk group) to 5.2 with the BBff genotype (Table 1)

**Table 1**

**Distribution of cases and controls among females with various VDR genotypes**

| VDR Phenotype | *BsmI* | Cases | Controls | Total | 95% Confidence Interval (CI) for Odds Ratio (OR) | | |
|---|---|---|---|---|---|---|---|
| | | | | | OR | Min | Max |
| BB | 1 | 36 | 92 | 128 | 1 | | |
| Bb | 2 | 111 | 198 | 309 | 1 4327 | 0.91 | 2.25 |
| bb | 3 | 101 | 160 | 261 | 1 6132 | 1.02 | 2.55 |
| | Totals | 248 | 450 | 698 | | | |
| | *FokI* | Cases | Controls | Total | OR | Min. | Max. |
| FF | 1 | 96 | 171 | 267 | 1 | | |
| Ff | 2 | 111 | 200 | 311 | 0 9886 | 0 70 | 1.39 |
| ff | 3 | 48 | 65 | 113 | 1 3154 | 0.84 | 2.06 |
| | Totals | 255 | 436 | 691 | | | |

The present invention will be more readily understood by referring to the following example which is given to illustrate the invention rather than to limit its scope.

**EXAMPLE 1**

**Polymorphism of the VDR Receptor Gene as a Marker for Breast Cancer Susceptibility**

In a case control study comparing 248 consecutive cases of breast cancer in women and 450 control women matched for age, two

polymorphism at the VDR gene, namely, the *BsmI* biallelic polymorphism located in the last intron of the gene (intron 8) and the *FokI* biallelic polymorphism located in the first exon were studied. Table 1 above presents the frequency of cases and controls in categories of *BsmI* and *FokI* genotypes independently with the corresponding odds ratio (OR) for breast cancer and the computed 95% confidence intervals (CI). The VDR gene *BsmI* alleles were identified as "B" for the absence of the restriction site and "b" for the presence of this site. This makes the BB genotype the least frequent VDR in our study population which is in keeping with other studies of this genetic marker in other populations. The *FokI* polymorphism was also identified as "F" when the restriction site was absent and "f" when it was present.

Table 1 shows that women with certain combination of VDR gene *BsmI* polymorphism (bb genotype) have a significant but slightly increased risk (OR of 1.6) of developing breast cancer as compared to the category with the smallest risk (BB genotype). However, these women represent about 35% of the general population. Taken alone, the *FokI* genotype was not associated with a significantly increased risk of breast cancer as the proportion of cases and control was roughly the same for each *FokI* genotype However, when the combined VDR genotypes (*BsmI* and *FokI*) were studied , a striking association with the disease was observed (Table 2). When considering the combined *BsmI* and *FokI* VDR genotypes, the odds ratio for breast cancer became extremely strong and significant when compared to the genotype combination that had the least number of cases (Table 2)

**Table 2**
**Distribution of cases and controls among females with various combined VDR genotypes**

| VDR geno-type | *BsmI* | Cases | Controls | Total | 95% Confidence Interval (CI) for Odds Ratio (OR) | | |
|---|---|---|---|---|---|---|---|
| | | | | | OR | Min. | Max. |
| BBFf | 1 | 11 | 43 | 54 | 1 | | |
| BBFF | 2 | 15 | 40 | 55 | 1 47 | 0 60 | 3 57 |
| BBff | 3 | 8 | 6 | 14 | 5.21 | 1 50 | 18 17 |
| BbFF | 4 | 47 | 75 | 122 | 2 45 | 1.15 | 5.22 |
| BbFf | 5 | 39 | 80 | 119 | 1 91 | 0 89 | 4.09 |
| Bbff | 6 | 22 | 30 | 52 | 2 87 | 1 21 | 6.78 |
| bbFF | 7 | 30 | 49 | 79 | 2.39 | 1 07 | 5.34 |
| bbFf | 8 | 51 | 69 | 120 | 2 89 | 1.36 | 6 15 |
| bbff | 9 | 18 | 26 | 44 | 2 71 | 1 11 | 6.62 |
| | Totals | 241 | 418 | 659 | | | |
| | *FokI* | Cases | Controls | Total | OR | Min. | Max. |
| BBFF + BBFf | 1 | 26 | 83 | 109 | 1 | | |
| Others | 2 | 215 | 335 | 550 | 2 05 | 1 28 | 3.29 |
| | Totals | 241 | 418 | 659 | | | |

* Standard Error
** 95% Confidence interval expressed with the maximum and minimum values

These odds ratios ranged from 1 (BBFf genotype) to 5.2 for the BBff genotype (95% CI . 2.3 to 11.5) while the majority of the other combined VDR genotypes had odds ratios for breast cancer between 2.4 and 3. When the number of total cases of breast cancer attributable to these two VDR polymorphisms was computed, it appeared that 45% of all cases were attributable to VDR gene polymorphism, as detected by the combined testing of the *BsmI* and the *FokI* genotypes (Table 3) This is thus the strongest susceptibility marker for breast cancer ever identified as it explains ten times more cases of the disease than the BRCA1 and BRCA2 genes combined. This

is due to the fact that the VDR genotypes associated with an increased susceptibility to breast cancer are the most frequent ones. Alternatively, the VDR BBFF and VDR BBFf genotypes could be seen as "protective". Together, these two latter genotypes represent about 18% of the genotypes found in the population.

**Table 3**

**Number of total cases of breast cancer attributable to VDR gene variation**

| Genotypes VDR *BsmI* | | | | |
|---|---|---|---|---|
| | Attributable risk | | | Total |
| Cases | 36 | 111 | 101 | 248 |
| Controls | 92 | 198 | 160 | 450 |
| | 128 | 309 | 261 | 698 |
| | | | vs 36/92 | vs 147/290 |
| | | Attributable cases | 71.9130435 | 19.8965517 |
| | | % | 0.28997195 | 0.08022803 |
| COMBINED *BsmI* & *FokI* genotypes | | | | |
| | | VDR B+F | | |
| | BBFF+BBFf | others | TOTAL | |
| Cases | 26 | 215 | 241 | |
| Controls | 83 | 335 | 418 | |
| Totals | 109 | 550 | 659 | |
| | | Attributable cases | 110.060241 | |
| | | % | 0 4566815 | |

Indeed, in the cohort studied the 18% of women with the less at risk VDR genotypes combinations (BBFF and BBFf) comprised only 10% of all breast cancer cases while the 82% of women with the most at risk VDR

genotype combinations had 90% of all breast cancer cases. In other words, as compared with the general population risk of breast cancer of 1:9 women, women with certain VDR genotypes had a risk of 1 17 (much lower) while the other group had a risk of 1:8 (larger). Thus, this novel genetic marker of breast cancer allows to identify a subgroup of women with a risk of breast cancer close to two times larger than the other subgroup.

Until now, no marker with such a large odds ratios had been reported for breast cancer. Furthermore, this genetic marker and polymorphisms in the VDR gene appears to play a very significant role in breast cancer susceptibility in women. This also points to alternative therapies for breast cancer aiming at mimicking the effects of specific "protective" VDR genotype combinations in women.

Although the present invention has been described hereinabove by way of preferred embodiments thereof, it can be modified, without departing from the spirit and nature of the subject invention as defined in the appended claims.

**SEQUENCE LISTING:**

| | | |
|---|---|---|
| | 5'-CAACCAAGAC TACAAGTACC GCGTCAGTGA-3' | SEQ ID NO:1 |
| | 5'-TATCGTGAGT AAGGCAGGAG AGGGAGACC-3' | SEQ ID NO:2 |
| 5 | 5'-AGCTGGCCCT GGCACTGACT CTGCTCT-3' | SEQ ID NO:3 |
| | 5'-ATGGAAACAC CTTGCTTCTT CTCCCTC-3' | SEQ ID NO:4 |